

Continuing medical education

How to interpret an overview: a meta-analysis of the relative efficacy and toxicity of *Pneumocystis carinii* prophylactic regimens

Philippa J Easterbrook

Introduction

The terms “overview”, “systematic review” and “meta-analysis” are frequently used interchangeably to describe the statistical technique of pooling results from two or more independent but similar studies, for the purpose of integrating the findings. Meta-analyses are becoming increasingly popular in clinical trials to help draw a global conclusion as to the usefulness of a drug or procedure, but also in observational studies to assess the contribution of a risk factor to a disease. The main objectives of meta-analyses of clinical trials are: (i) to increase the statistical power for primary endpoints and in subgroups; (ii) to provide more stable estimates of a treatment effect; (iii) to resolve uncertainties when studies disagree; (iv) to aid in interpreting the generalisability of results; and (v) to conduct subgroup analyses for which individual trials have far too low a statistical power. The main difficulty in integrating the results of various studies stems from their often diverse design and analytical methods.

Background to clinical problem

One of the landmark developments in the management of HIV infection has been the widespread introduction of primary prophylaxis against *Pneumocystis carinii* pneumonia (PCP)—the commonest AIDS opportunistic infection in industrialised countries.¹ Most clinicians now follow the recommendations to initiate primary prophylaxis when the patients' CD4 cell count falls below $200 \times 10^6/l$.² As a result, the frequency of PCP and its associated morbidity and mortality have decreased markedly in Western countries. Several prophylactic regimens are available, but the choice of agent can be problematic. In general, trimethoprim-sulphamethoxazole (co-trimoxazole) (TMP-SMX) is recommended as the first line agent at a dose of one double strength

tablet daily. However, while oral regimens are effective and inexpensive, a significant proportion of patients develop serious adverse reactions or fail to tolerate therapy. Alternative agents include aerosolised pentamidine or dapsone, but while pentamidine is better tolerated than TMP-SMX, it lacks the additional anti-toxoplasma activity of the oral regimens. Regimens of varying dose and frequency have also been proposed,¹ but the optimal dosing regimen remains unclear.

Many randomised trials have been conducted over the past 8 years to identify the optimal prophylactic regimen. However, because of the relatively small number of patients enrolled in each study, the results of the individuals trials are often difficult to interpret, and few have examined the relative efficacy and toxicity of different dosing regimens of PCP prophylaxis.³

A recent meta-analysis of 35 trials was conducted to evaluate the efficacy and tolerability of different strategies of PCP prophylaxis.⁴ In this article, I will use this example to demonstrate the key points to look for in assessing the validity and clinical relevance of a meta-analysis, as summarised in table 1. A further guide on interpreting overviews is available from Oxman and the Evidence-Based Medicine Working Group.⁵

What are the results of the meta-analysis?

Meta-analyses differ from careful literature reviews in that they create and analyse at least one numeric summary for each primary study. A variety of statistical techniques have been developed for combining the results from multiple trials, but they all share the same basic approach: comparing the observed findings from the available trials to the findings that would be expected if there was no effect of treatment. One of the most widely used techniques is the pooled log odds ratio—the Mantel-Haenszel method which combines separate two by two tables. This method has several useful properties: it compares each treatment with its own control, and weights each study according to its sample size. The method assumes that trials asking similar questions should (except for the play of chance) yield answers that point in the same direction, regardless of the nature of the study population. In performing a meta-analysis, there is no intention of trying to generate exact quantitative estimates of percentage risk reductions in some precisely defined population of patients, but simply to determine whether or not a given

HIV Epidemiology Unit, Chelsea and Westminster Hospital, 369 Fulham Road, London SW10 9TH, UK
P J Easterbrook

Accepted for publication
8 October 1996

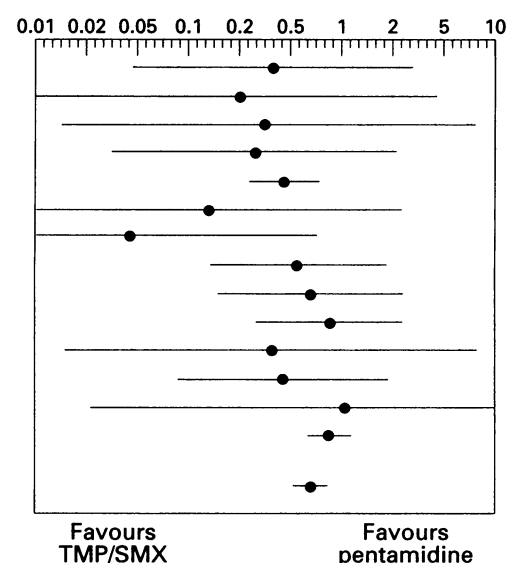
Table 1 Readers' guide on how to assess a meta-analysis

What are the results of the meta-analysis?
What are the overall results of the meta-analysis?
How precise were the results?
Are the results of the meta-analysis valid?
Did the meta-analysis address a focused clinical questions?
Were the criteria used to select articles for inclusion appropriate?
Is it unlikely that important relevant studies were missed?
Was the validity of the included studies appraised?
Were assessments of the studies reproducible?
Were the results similar from study to study?
Will the results help me in caring for my patients?
How generalisable are the results to patients in my care?
Were all clinically important outcomes, including side effects and costs, considered?

Adapted from Oxman *et al.*⁵

*Risk ratios for failure of *Pneumocystis carinii* prophylaxis in studies comparing trimethoprim-sulphamethoxazole (TMP-SMX) with aerosolised pentamidine. Reproduced from Ioannidis et al.⁴*

Investigators	Year	No of patients
Antinori et al	1991	45
Nielsen et al	1991	75
Rozenbaum et al	1991	55
Antinori et al	1992	134
Hardy et al	1992	310
Michelet et al	1992	108
Shneider et al	1992	213
Smith et al	1992	53
Mallolas et al	1993	215
Rizzardi et al	1993	200
Stellini et al	1993	41
May et al	1994	214
Toconetti et al	1994	30
Bozzette et al	1995	554
Overall		2248



treatment, tested in a wide range of trials, has any effect on outcome.

Meta-analyses are often presented visually by charting their odds ratios and associated confidence intervals. The figure shows the risk ratios for failure of PCP prophylaxis in studies comparing TMP-SMX with aerosolised pentamidine.

For each trial, the observed odds ratio is plotted as a black circle, with its 95% confidence interval as a horizontal line. The bottom circle represents the odds ratio and 95% confidence interval for the overview of all the individual trials. The solid vertical line represents an odds ratio of unity. The closer the assembled trials cluster around 1 (that is, no effect), the less likely is any treatment effect.

The meta-analysis conducted by Ioannidis and colleagues pooled data from 35 randomised trials of PCP prophylaxis involving 6538 patients.⁴ The analysis confirmed that in patients who received any form of PCP prophylaxis compared with placebo, episodes of PCP (and deaths due to PCP) were significantly reduced (risk ratio = 0.32 and 0.83 respectively; relative risk reduction = 1 - relative risk, that is, 1 - 0.32 = 68%; 95% CI 54% to 77%). In the comparison of different prophylactic regimens, TMX-SMX was superior to aerosolised pentamidine with a 42% (CI 24% to 55%) overall reduction in prophylaxis failures. Regardless of dose (range of 2 double strength tablets per day to 1 double strength tablet three times per week), TMX-SMX was almost universally effective for patients who tol-

erated it (table 2). There was a non-significant trend towards fewer treatment failures with TMX-SMX compared with dapsone based regimens (relative risk reduction = 39%; 95% CI 10% to 66%). Aerosolised pentamidine and dapsone based regimens were equivalent in effectiveness, but prophylaxis failures were halved when the dose of aerosolised pentamidine was doubled to 300 mg monthly (table 2). For dapsone, the multivariate model predicted that among 100 patients given 100 mg daily instead of twice a week for 1 year, seven fewer patients would develop PCP (2.7% v 10.0% in 1 year), but 17 more would have a significant adverse event (29% v 12% in 1 year). TMX-SMX or dapsone regimens also reduced the number of cases of toxoplasmosis by 33% (CI 12% to 50%) compared with aerosolised pentamidine, based on an intention to treat analysis. Compared with aerosolised pentamidine, however, oral regimens were five times more likely overall to be discontinued because of toxic reactions. The final multivariate model predicted that the risk for discontinuing TMX-SMX because of side effects would decrease from 23.2% to 14.5%; corresponding to a risk reduction of 43% (CI 30% to 54%) when one double strength tablet was given three times per week instead of daily. No difference existed in overall mortality among the various regimens (table 3).

Are the results of the study valid?

Since meta-analysis involves a retrospective look at data already collected, it is important to make the process rigorous and well defined to prevent opportunities for bias to distort the results. As with any research study, the questions to be answered, the criteria for inclusion in the study, and the methods to be used should be established beforehand (table 3).

DID THE OVERVIEW ADDRESS A FOCUSED CLINICAL QUESTION?

In order to assess whether the findings of the meta-analysis are relevant to the reader's clinical practice, the question for each meta-analysis should be explicitly stated and clinically relevant.

Table 2 Relative merits of different doses of *Pneumocystis carinii* pneumonia prophylactic regimens

Prophylaxis	Dose	Prophylaxis failure rate (%)		Severe toxic effects (in 1 year)
		Primary (in 1 year)	Secondary (in 1 year)	
Trimethoprim-sulphamethoxazole	Double strength tablet/day	0.5	0.5	23.2†
	Double strength tablet thrice/week	1.8	0*	14.5†
Dapsone	100 mg/day	2.7	No data	29
	100 mg twice/week	10.0	1.7*	12
Aerosolised pentamidine	300 mg/month	7.5	1.7	2.5
	300 mg twice/month	4.9*	< 1*	1*

Adapted from Table 6, Ioannidis et al.⁴

*Few data available.

†Data for US patients only.

Table 3 Pooled risk ratios for *Pneumocystis* prophylaxis failures, mortality, and toxoplasmosis for comparisons of different prophylactic regimens

Comparison	PC events	All deaths	Toxoplasmosis events
Any prophylaxis <i>v</i> placebo	0.32 (0.23–0.46)	0.83 (0.63–1.09)	
Primary prophylaxis	0.39 (0.27–0.55)	0.87 (0.60–1.25)	
Secondary prophylaxis	0.16 (0.08–0.35)	0.84 (0.33–2.11)	
Oral regimens <i>v</i> aerosolised pentamidine	0.73 (0.59–0.91)	0.99 (0.89–1.09)	0.67 (0.50–0.88) (17; 3882)
Trimethoprim-sulphamethoxazole <i>v</i> aerosolised pentamidine	0.58 (0.45–0.75)	0.99 (0.80–1.22)	0.78 (0.52–1.16) (13; 2158)
Dapsone <i>v</i> aerosolised pentamidine	0.93 (0.72–1.19)	0.98 (0.86–1.12)	0.61 (0.43–0.87) (8; 2137)
Trimethoprim-sulphamethoxazole <i>v</i> dapsone	0.61 (0.34–1.10)	0.95 (0.82–1.11)	1.26 (0.68–2.34) (6; 1293)

Adapted from Table 2, Ioannidis *et al.*⁴

All analyses are based on intention to treat data.

In the meta-analysis under review, the authors state clearly that the objective is to evaluate the efficacy and tolerability of different strategies of PCP prophylaxis in HIV infected patients, and so make recommendations about the optimal strategy for prophylaxis.

WERE THE CRITERIA USED TO SELECT ARTICLES FOR INCLUSION APPROPRIATE?

In determining whether the investigators reviewed the research studies relevant to the main clinical question, the reader needs to know that an established protocol for selection of studies for inclusion was developed at the start of the meta-analysis. These criteria should include the nature of the patient population, treatments, dosing regimen, and relevant outcomes. For example, in the meta-analysis under review, the investigators could have confined themselves to studies of oral prophylaxis regimens only, to a particular dosing regimen, or to either primary prophylaxis (that is, among patients without a prior history of PCP), or secondary prophylaxis studies (that is, among patients who already had an episode of PCP). Some investigators may choose to include observational studies in addition to randomised trials, although the lack of randomisation in such studies can introduce substantial bias favouring the new treatment and lead to spurious results.⁶ To avoid bias in selecting and rejecting papers, the decision to include a paper should be made by blinding the selection of papers (that is, by only looking at its methods and not at its results or the authors' location or the journal). The investigators also need to state the methodological standards used to select articles, and the data should be extracted in duplicate. For clinical trials these standards will be similar to those employed to assess the validity of the evaluation of original reports of trials.⁷

This meta-analysis gave clear evidence that the study was conducted according to a predetermined protocol of selection criteria. The authors considered data from all randomised studies that used TMP-SMX, dapsone based regimens, or aerosolised pentamidine in adult patients, and that compared different prophylaxis regimens directly with each other or with placebo. Both primary and secondary prophylaxis trials were included. From each trial, the investigators assessed the number of PCP episodes, PCP related deaths, toxoplasmosis events, deaths, adverse events and discontinuations.

IS IT UNLIKELY THAT IMPORTANT RELEVANT STUDIES WERE MISSED?

The investigators should provide details of their search procedures in order to demonstrate that they have conducted a thorough search for all studies that meet their stated inclusion criteria. The meta-analysis should include as many relevant randomised clinical trials as possible, so as not to exclude important data that might significantly alter the results. This should ideally involve searching one or more computerised bibliographic databases, such as MEDLINE or EMBASE, (using specific terms such as "random allocation", "randomised", "double blind", "controlled"). However, it is insufficient to rely solely on computer searches of the literature since they may yield less than two thirds of relevant trials. A computer search can be supplemented by consulting *Current Contents*, reviews, textbooks, manual journal review, checking the reference list of the articles retrieved, and conference abstracts. Finally, the use of personal contacts in the field may be important to identify other published studies that may have been missed, as well as to help in the identification of unpublished studies. There is concern that omission of unpublished studies from a meta-analysis, creates a selection bias towards studies with favourable results due to a "publication bias", whereby studies with positive results are more likely to be published than those with negative results.⁸ However, the importance of including unpublished studies remains controversial, since such studies may remain unpublished for good reason, such as major methodological flaws. A simple method has been proposed for calculating the number of unpublished negative studies required to refute the published evidence, which may serve as a useful measure of the strength of the published evidence.⁹

In this meta-analysis, randomised controlled trials in the English language literature were identified through MEDLARS, in addition to abstracts from the key international AIDS conferences. As a further crosscheck, the non-English language literature and the reference list from the trials identified through the MEDLARS search were screened to ensure all relevant trials had been identified. The investigators, however, did not use personal contacts and unpublished studies were not included. No mention was made of the potential problem of publication bias, although this is unlikely to be substantial,

since most sizeable trials on this important clinical topic are likely to be published, or at least presented in abstract form.

WAS THE VALIDITY OF THE INCLUDED STUDIES APPRAISED?

The validity of a meta-analysis is largely determined by the methodological rigour and scientific quality of the individual contributing trials. Differences in study methods and small sample size may explain important differences in findings. Several approaches to assessing the quality of clinical trials have been proposed. Some of these involve applying checklists and scoring the trial on each item, while others focus on a few key features in the design and analysis of a clinical trial.^{10 11} Such features include the randomisation process, the measurement of patient compliance, the blinding of patients and observers, the statistical analysis, and the handling of withdrawals. Of these, the most important in terms of impact on the validity of results is the method of treatment assignment in the primary study. It has been shown that trials using historical controls tend to overestimate the effectiveness of therapeutic and preventive interventions than trials where treatment assignment is random.⁶

The possibility of bias in patient inclusion may be suggested by imbalances in the total number of patients in the different treatment groups, or trials by imbalances within certain strata. In order to ensure that the analysis is unbiased, it is important to check that outcome was compared among all those allocated treatment with all those allocated control (that is, "intention to treat" analysis), even if some of the patients did not actually receive the allocated treatment. Finally, in the assessment of any meta-analysis, it is important that the numbers of withdrawals and losses to follow up in the individual trials are small, so that biases introduced are minimal. Biases may be introduced by losses to follow up which may be both outcome and/or treatment dependent. Evidence of exclusions after randomisation may be suggested by imbalances in the number of patients reported to have been allocated to the different treatment groups, or by differences between the treatment groups in the frequency, duration or type of follow up.

In this meta-analysis, no formal quality checks were applied to the individual trials. However, the investigators set two key methodological requirements for inclusion: randomised treatment allocation and analysis of data on an "intention to treat basis".

WERE ASSESSMENTS OF STUDIES REPRODUCIBLE?

The application of clear selection criteria for the inclusion of trials in the meta-analysis is an important step in ensuring the validity of the meta-analysis. However, there is scope for subjective interpretation of these criteria, and different observers may disagree. One way to guard against this type of bias is to have two or more people involved in selecting, reviewing and abstracting data from each trial, and then to measure the interobserver agreement.

It is not possible in the meta-analysis under consideration to assess whether the assessments were reproducible. The investigators do not state who was responsible for applying the selection criteria to the identified articles. However, the clarity of the criteria reduces the likelihood of significant selection bias by one investigator.

WERE RESULTS SIMILAR FROM STUDY TO STUDY?

A major issue in pooling data is whether the results of the separate trials can be meaningfully combined. Even with well defined selection criteria, the trials included in a meta-analysis are likely to differ in patient characteristics, study designs, interventions and outcomes. However, an underlying assumption of the meta-analytic approach is that any differences in results across trials are due to chance alone, and that if different trials are addressing the same question, then there is going to be some tendency for the answers to come out in the same direction. This tendency may be obscured in individual trials, or in some cases reversed, by the play of chance, but when results are combined, these random errors cancel out. Nevertheless, before combining data, an initial assessment needs to be made as to whether the differences between the trials are so great as to render the pooling of data inappropriate. Although determining the comparability of trials requires mainly common sense, as a first step, a graphical display of the results of individual studies is helpful. Formal statistical tests of "heterogeneity" or between study variability can then be used to compare observed variation to what would be expected simply due to sampling. However, such tests need to be interpreted cautiously. For example, a non-significant test of heterogeneity may obscure the presence of clinically important differences between study results. Conversely, even when there is evidence of significant heterogeneity, the pooled summary odds ratio may still provide the best estimate of the likely impact of an intervention.

Non-uniformity of study outcomes may reflect the fact that the treatment effect varies according to a particular study characteristic, such as age, stage of disease or treatment dose. In this situation, such techniques as stratification of the trials by selected patient characteristics, and reporting of separate results for each stratum will reduce the problem of heterogeneity. Regression techniques can also be used to identify various factors such as dose contributing to differences in the treatment effect.

All readers should look for some discussion in the meta-analysis report on the issue of homogeneity. Any substantial differences in the primary studies should be noted, and there should be some discussion of how these differences affected the conclusions. This meta-analysis fulfilled these requirements, by presenting the results in both a tabular and graphical form, and reporting a formal test of heterogeneity. In addition, the use of regression analyses allowed an evaluation of the

impact of different dosing regimens on treatment effect.

Will the results help me in caring for my patient?

HOW GENERALISABLE ARE THE RESULTS TO PATIENTS IN MY CARE?

In this meta-analysis, data were collected from more than 6000 HIV positive patients enrolled from different risk groups and different geographical regions. The consistency of results across such a diverse range of patients implies that the findings are widely generalisable. However, many of these trials were conducted before the current era of combination antiretroviral therapy, and may therefore both overestimate prophylaxis failure and potentially underestimate the true rate of adverse events as a result of drug interactions. It is also unclear how well these findings can be extrapolated to secondary prophylaxis regimens, since none of the trials fully examined the efficacy of low dose regimens in secondary prophylaxis.

WERE ALL CLINICALLY IMPORTANT OUTCOMES, INCLUDING SIDE EFFECTS AND COSTS CONSIDERED?

There are several well known examples of where benefit was derived from an intervention in one outcome, but showed harm in another. For example, certain trials of cholesterol lowering drugs show a benefit on deaths from coronary heart diseases, but appear to be associated with a greater likelihood of death from causes other than heart disease.

In the meta-analysis, all relevant outcomes were considered including PCP episodes, PCP related deaths, overall mortality, different side effects and rate of discontinuation of medication. An additional evaluation of the economic impact on the different prophylactic strategies would have been helpful.

In summary, this meta-analysis meets all the important validity criteria, with only a few limitations. Direct comparisons of different regimens were few or did not exist at all, and as a result conclusions were based on indirect comparisons of treatment arms across different clinical trials. Although adjustment was made in the regression analysis for some important confounders or effect modifiers of treatment effect across trials such as country, and dose, the effect of variables such as CD4 cell count and use of antiretroviral agents with potential drug interactions could not be examined with most studies. Finally, the meta-analysis did

not include an explicit assessment of the validity of the individual trials, and from the graphical and tabular presentation of the results, there appears to be clinical heterogeneity across the studies.

The results of this meta-analysis suggest that low dose TMX-SMP (one double strength tablet three times a week or one single strength tablet each day) should be adopted as the standard first line regimen in primary prophylaxis. This approach, combined with the use of more refined protocols for desensitisation, may reduce the number of patients who fail to tolerate this drug. For those patients who remain intolerant, the results also suggest that a dose increase in aerosolised pentamidine from 300 mg per month to 300 mg twice per month would result in a 50% decrease in prophylaxis failures. In contrast, although lower doses of dapsone were better tolerated than higher doses, some efficacy was lost. The dose of dapsone should therefore be determined on a case by case basis according to baseline haematological status and use of interacting drugs, such as didanosine. Finally, questions remain about the contribution of desensitisation to TMP-SMX and the role of other potentially useful regimens such as pyrimethamine-sulphadoxine (Fansidar), intramuscular pentamidine, clindamycin-primaquine and atovaquone.

- 1 Gallant JE, Moore RD, Chaisson RE. Prophylaxis for opportunistic infections in patients with HIV infection. *Ann Intern Med* 1994;120:932-4.
- 2 US Public Health Service Task Force on antipneumocystis prophylaxis in patients with human immunodeficiency virus infection. Recommendations for prophylaxis against *Pneumocystis carinii* pneumonia for persons infected with human immunodeficiency virus infection. *J Acquir Immun Defic Syndr* 1993;6:46-55.
- 3 Der Simonian R, Laird NM. Meta-analysis in clinical trials. *Control Clin Trials* 1986;7:177-88.
- 4 Ioannidis JP, Cappelleri JC, Skolnick PR, Lau J, Sacks HS. A meta-analysis of the relative efficacy and toxicity of *Pneumocystis carinii* prophylactic regimens. *Arch Intern Med* 1996;156:177-88.
- 5 Oxman AD, Cook DJ, Guyatt GH, for the Evidence-Based Medicine Working Group. Users' Guide to the Medical Literature. VI. How to Use and Overview. *JAMA* 1994;272:1367-71.
- 6 Chalmers TC, Celano P, Sacks HS, Smith H. Bias in treatment assignment in controlled trials. *N Engl J Med* 1983;309:1358-61.
- 7 Guyatt GH, Sackett DL, Cook DJ. Users guide to the medical literature. II. How to use an article about therapy or prevention. A. Are the results of the study valid. *JAMA* 1993;270:2598-601.
- 8 Easterbrook PJ, Berlin JA, Gopalan R, Matthews DR. Publication bias in clinical research. *Lancet* 1991;337:867-72.
- 9 Begg CB. A measure to aid the interpretation of published clinical trials. *Stat Med* 1985;4:1-9.
- 10 Chalmers TC, Smith H, Blackburn B, Silverman B, Schroeder B, Reitman D, et al. A method for assessing the quality of a randomized control trial. *Control Clin Trials* 1981;2:31-49.
- 11 L'Abbe KA, Detsky AS, O'Rourke K. Meta-analysis in clinical research. *Ann Intern Med* 1987;107:224-33.